Instructions for Assembling Datasets for Demand Estimation

February 19, 2020

Assemble Monthly Sales Dataset The first task is to assemble a dataset with sales (quantity and average price) by month and designated market area (DMA). Vivek has put together some R code (/projects/b1048/gillanes/Mergers/aggregation_code.R) that provides a general framework for this step. Feel free to use this, as well as the code Yintian has put together, as guidance to create a script to pull the data for different years and product codes.

In order to standardize across different product sizes, there are a few changes to how we'll be calculating quantities and prices. Before aggregating to the DMA/month level, you'll first have to create two variables: total sales and total volume (in size1_units) for each observation in the movement dataset.

$$\text{Total Sales} = \frac{\text{price}}{\text{prmult}} \times \text{units}$$

$$\text{Total Volume} = \text{units} \times \text{size1\_amount} \times \text{multi}$$

A few of the variables from above (size1_amount and multi) come from the products.tsv file, while the others are contained in the movement files. Next, aggregate to the DMA/month level, computing the sum of sales and volume, as well as the average price per size1_unit.

Ideally, we'll be able to run this from the command line with the same inputs as are required in aggregation_code.R:

1. Vector of years (line 54): This vector should include all years in the raw data folder, as they've been chosen to span the period around the merger. For both the Mars/Wrigley merger and Miller/Coors joint venture, the relevant years are 2006, 2007, 2008, and 2009.

2. Nielsen product code (line 55): The relevant codes will vary by merger (e.g., 5001 for beer, 0503 for candy, and 0505 for gum). For Mars/Wrigley, we'll want both 0503 and 0505.

The script creates a dataset of the form *product code*.csv, which will be used as input for demand estimation once we augment it with product characteristics and calculate market shares.

Note that you'll likely have to run this code as a batch job due to memory constraints. This will be a good way to familiarize yourself with the submission scripts and Quest more generally.

Identify Top 100 Brands by Market Share The next step is to begin the process of constructing the ownership matrix by mapping brands in the Nielsen data to parent companies.

First, use *product code*.csv and products.tsv to create a dataset that lists, by month, the brands with the top 100 market shares nationally. Here, market shares should be calculated using the volume measure described in the first section.

We'll want this code to be generalizable, taking a product code (or product codes for mergers that span multiple)[1] as an input and returning a dataset with the union of the sets of brands that appear in the top 100 market shares in any given month. To identify brands, you'll have to merge products.tsv onto *product code*.csv using the UPC code and grab brand_code_uc and brand_descr.

Using the list of top brands, match each brand to a parent company using outside sources. The best way to complete this task will depend on the merger.

- For some mergers (namely, those occuring in 2010 or later), Euromonitor Passport will be a useful source. This platform lists market shares by brand/parent company combination and year. Brand/-parent company pairs that are active in a given year have a positive brand share. To determine the ownership structure pre- and post-merger, examine brand mappings in the year directly before the merger and the year directly after the merger.
  - To access Euromonitor, begin at https://libguides.northwestern.edu/businessatozlist/p and select "Passport Database (Euromonitor)." From there, log in using your Northwestern NetID and password.
  - Then, search "Brand Shares *Industry* in USA" and click on the first result in the drop down menu. The industries in the Euromonitor dataset do not match product types in the Nielsen data, but the relevant categories should be evident. For example, the industry for the Mars/Wrigley merger is "Packaged Food."
- For mergers outside of the Euromonitor dataset, other sources must be used. For the merging parties, SEC press releases could be useful. For example, the release for the Mars/Wrigley merger references brand portfolios for both merging parties.[2] For non-merging competitors, some manual search may be required. However, because these are brands with substantial U.S. sales, it should be more straightforward to find the associated parent companies than it was in the SABMiller/AB InBev merger.

[1] For example, the Wrigley/Mars merger spans both 0503 and 0505.
[2] https://www.sec.gov/Archives/edgar/data/108601/000134100408000776/wwy_ex99-2.htm